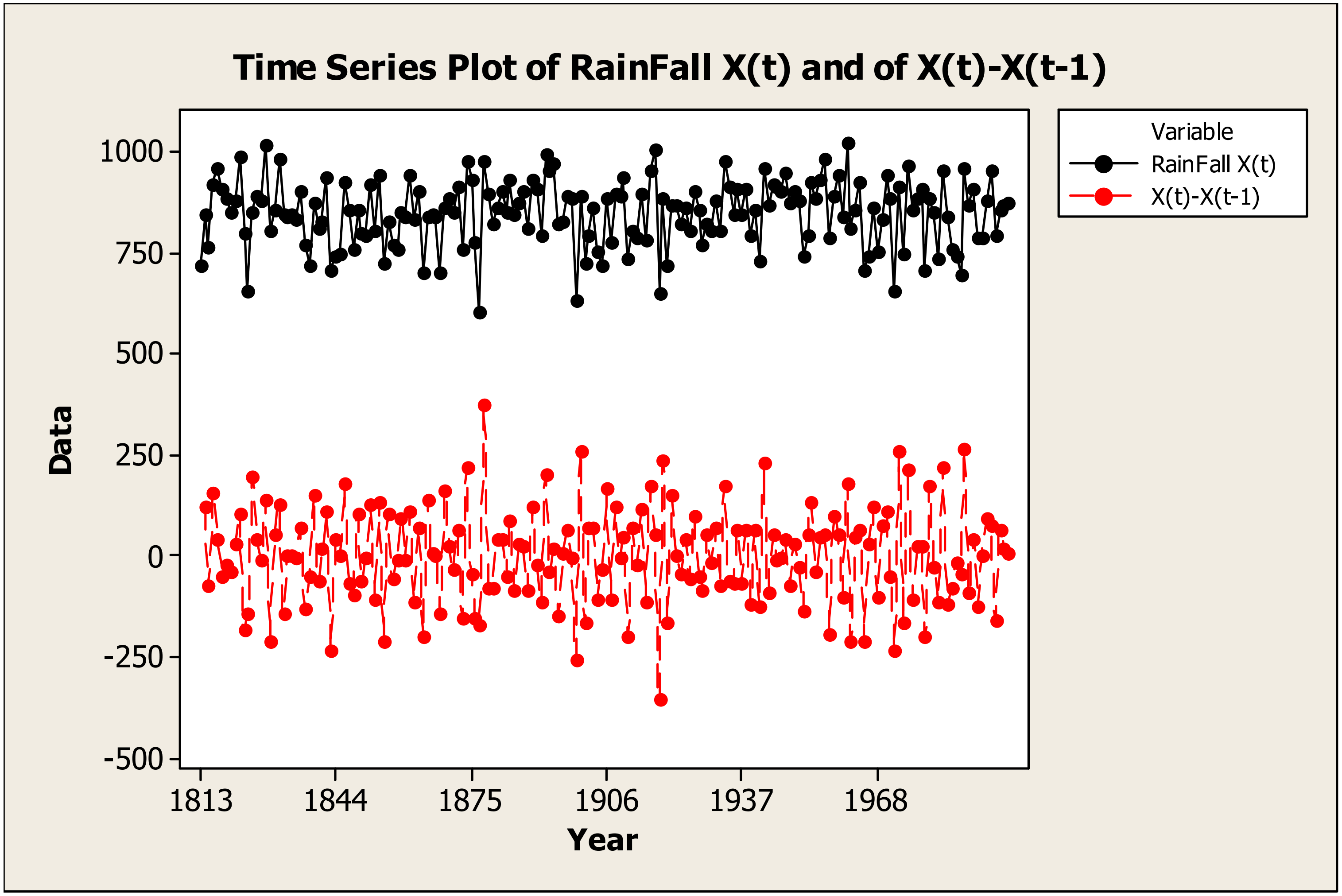
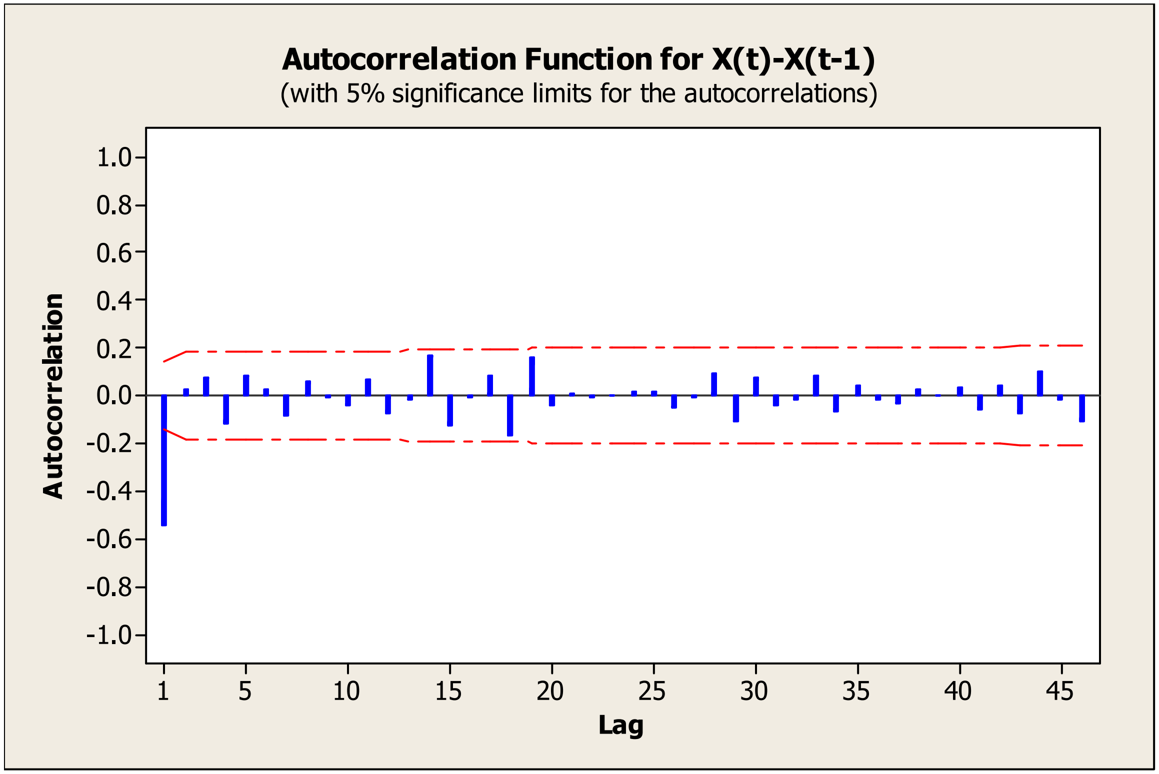
3.1 India Monsoon Data



- the data X_t show some local increasing or decreasing trends
- the variability of the data looks constant
- there are some exceptional years of very low rainfall or very high rainfall

- the detrended data by the difference method give a new series $Y_t = X_t - X_{t-1}$ which represents difference in rainfall of two consecutive years (positive indicates increase, negative indicates decrease in rainfall compared to the previous year)
- the new series oscillates about zero, local trends are removed
- the variance seems to be constant



- The sample ACF of the changes in rainfall data cuts off after lag 1 and the rest of the values "randomly" oscillate about zero. This suggests an MA(1) model.
- The sample autocorrelation at lag 1 is negative, equal to -0.54. It means that the moving average parameter θ is negative as

$$\rho(1) = \frac{\theta}{1+\theta^2}$$

The **ARIMA(0,0,1) ≡ MA(1)** option in Minitab gives the following output:

```
Final Estimates of Parameters

Type       Coef  SE Coef         T      P
MA   1   0.9980   0.0000  32278.19  0.000
```

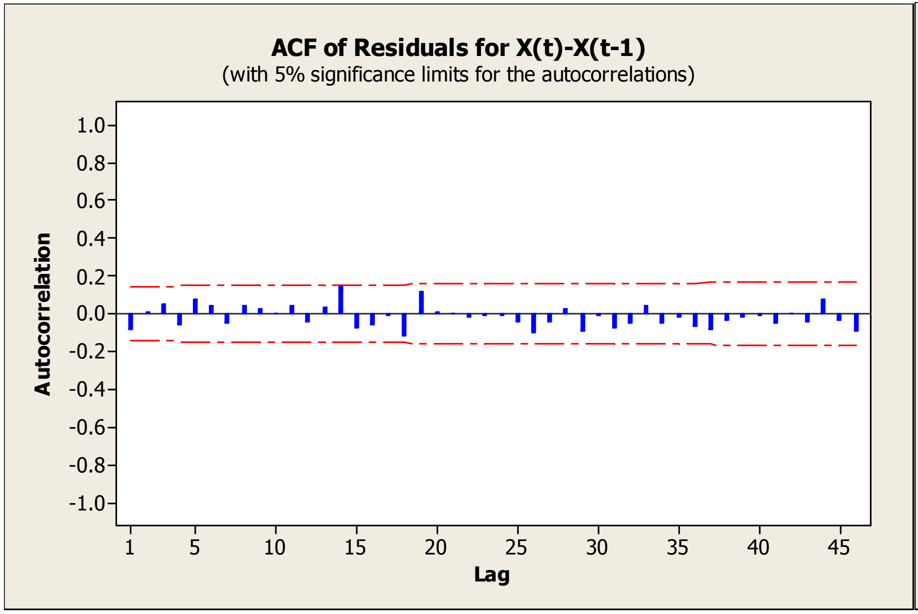
Here the estimate of coefficient θ is positive – this is because Minitab uses -θ in place of θ in our notation.

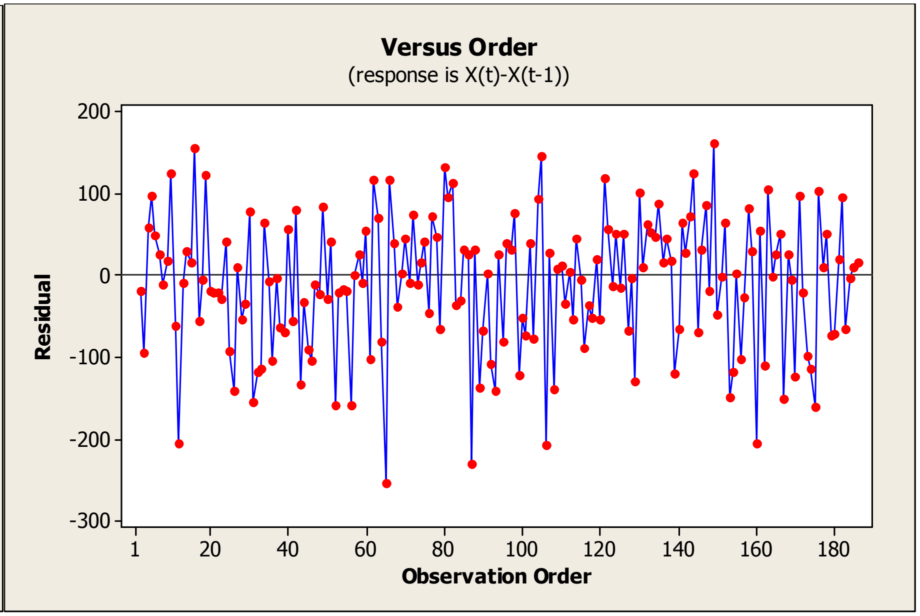
```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            12     24     36     48
Chi-Square    5.6   18.3   27.3   36.1
DF             11     23     35     47
P-Value     0.900  0.740  0.821  0.875
```

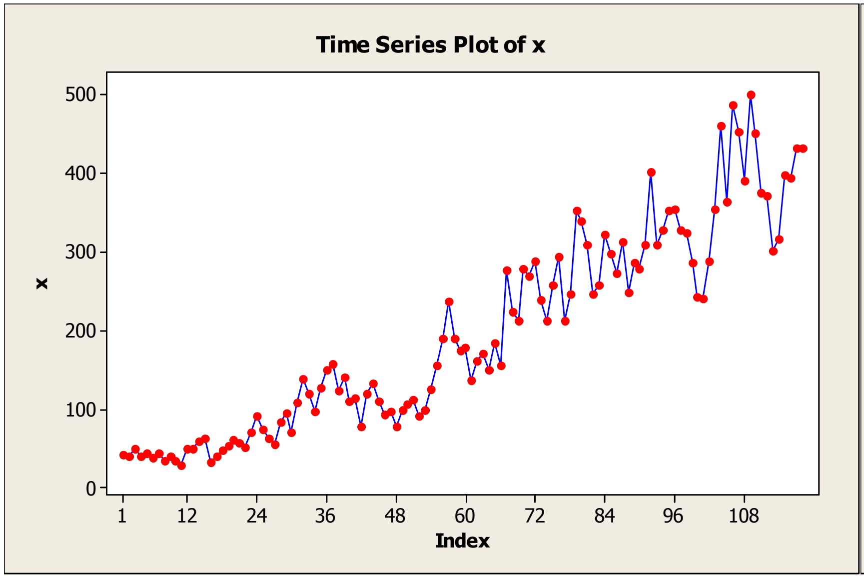
The T-test tells us about the significance of the MA parameter θ while the Ljung-Box Q (LBQ) statistic tests the null hypothesis that the autocorrelations of the residuals for all lags up to lag k are equal to zero.

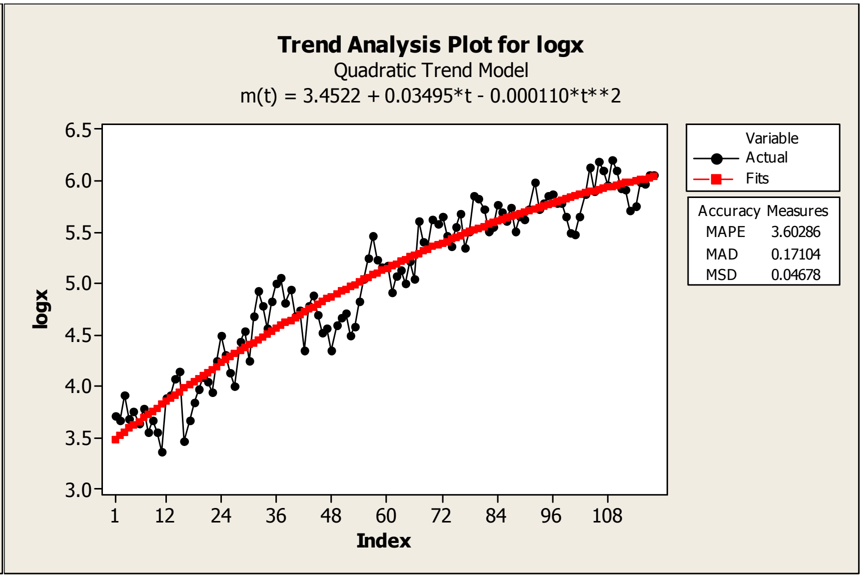
The results indicate that MA(1): $y_t = z_t - 0.998\, z_{t-1}$, where z_t is the realization of $Z_t \sim WN(0,\sigma^2)$ and y_t is the differenced series $x_t - x_{t-1}$, fits the data well. The two graphs below confirm the numerical results about the White Noise Z_t, which looks like a stationary, uncorrelated series.



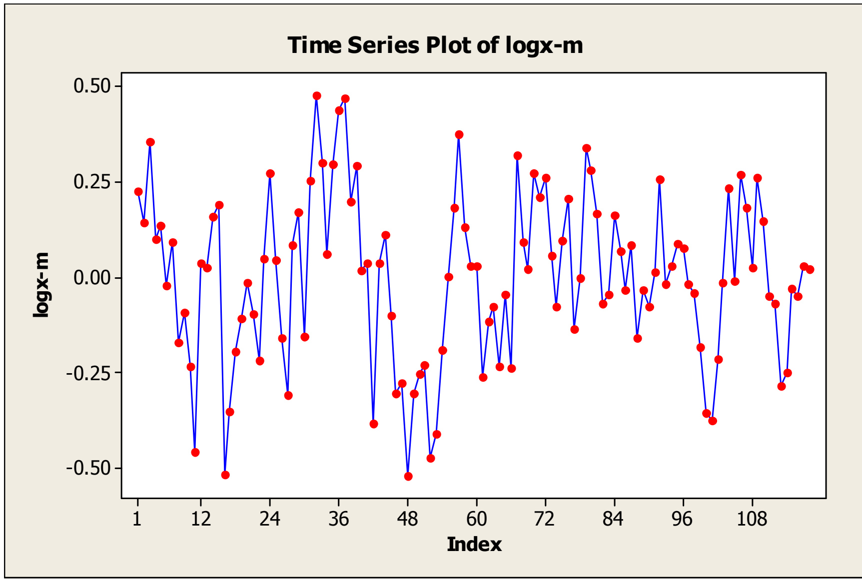


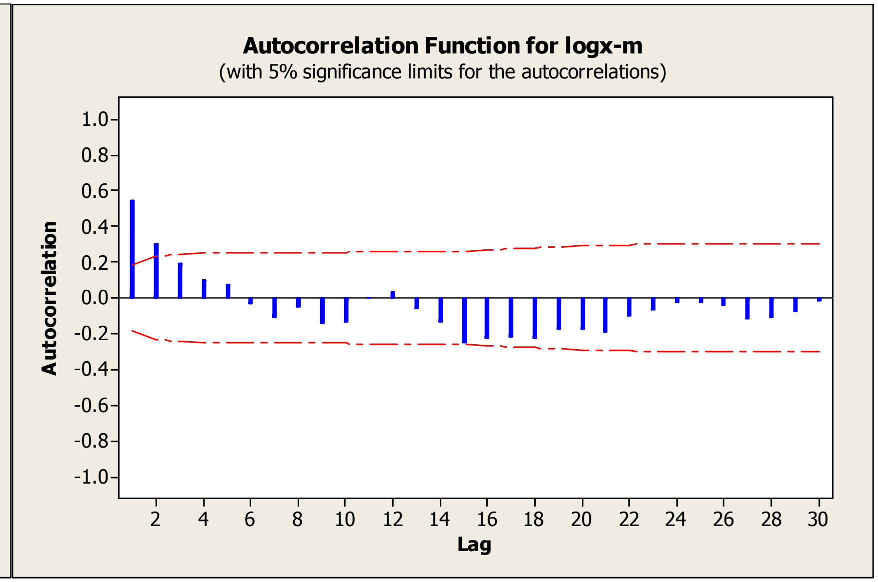
3.2 Boston Crime Data





- There is an increasing trend and also increasing variability in the data.
- A transformation is necessary for stabilizing the variance.
- The log transformed data do not show an increasing variability.
- The quadratic model fits the trend m(t) well.





- The detrended data are scattered about zero, but there are some local trends, which suggest possible correlations.

- The sample ACF indicates an AR or an ARMA model.

ARIMA(1,0,1) ≡ ARMA(1,1) model fit gives the following output:

```
Final Estimates of Parameters

Type       Coef  SE Coef     T      P
AR   1  0.5566   0.1401  3.97  0.000
MA   1  0.0095   0.1686  0.06  0.955
```

This means that the AR parameter is significantly different from zero, while the MA parameter is not. An AR(1) might be a better choice than ARMA(1,1).

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12     24     36     48
Chi-Square    11.6   22.0   36.7   49.7
DF              10     22     34     46
P-Value      0.314  0.457  0.343  0.328
```

The LBQ statistics show that the residuals might indeed be the required white noise values. The p-values are too large to reject the hypotheses of non-significant correlations of groups of residuals.

ARIMA(1,0,0) ≡ AR(1) model fit gives the following output:

```
Final Estimates of Parameters

Type       Coef  SE Coef      T      P
AR   1  0.5499   0.0772  7.13  0.000



Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            12     24     36     48
Chi-Square   11.5   22.1   36.7   49.6
DF             11     23     35     47
P-Value     0.399  0.515  0.388  0.369
```

The AR(1) model fit indicates a highly significant parameter φ with its estimate $\hat{\varphi} = 0.5499$.
The LBQ statistics, the graphs of the sample ACF and the plot of the residuals versus order indicate stationary uncorrelated series.

The model fit is $y_t = 0.5499\, y_{t-1} + z_t$, where z_t is the realization of a $WN(0,\sigma^2)$.